**IEEE SPMB 2023: Change Summary**

**Submission Type:**  Abstract

**Abstract No. A003:** An Evaluation of SMILE on the TUSZ Corpus

**Reviewer 1:**

“The abstract could be improved by providing more details about the evaluation methodology. For example, it would be helpful to know the ML algorithms”.

The abstract was modified significantly to help address these kinds of concerns. The third paragraph specifically addresses this concern.

“and how the data was split into training, development, and evaluation sets.”

The database is referenced here:

“In this abstract, we present an evaluation of SMILE on the well-known Temple University Hospital Seizure Corpus (TUSZ) [3].”

The database construction is described in this publication. Since this is an abstract, these kinds of details do not need detailed discussion. This corpus has been in use for many years.

**Reviewer 2:**

“Clearly state the contribution of this abstract.”

We addresss this in the second paragraph of the abstract.

“Provide a reference about NEDC ResNet-18 system.”

Done – pleae see the third paragraph.

“It would be easier to read if there is a mention somewhere that ‘FP rate’ stands for ‘false positive rate’.”

FP is defined in the second paragraph. We made the terminology consistent. False alarm rate is no longer used. FP Rate is also discussed in reference [6].

**Reviewer 3:**

“NEDC ResNet 18 was not defined at all in the text or even referenced”

Paragraph 3 has been created to describe the NEDC ResNet-18 system. A reference has also been added.

“In one part of the paper, it is mentioned ‘… The solution to this is to utilize the Linux workload manager, Slurm’. I don’t think the solution to computational expenses is the use of a free software like Slurm. Slurm is just a software that can help in easily operating and managing the workload on a HPC. The specifications of the HPC ( CPU cluster as mentioned) should be clearly defined.’

They text was modified to address this concern:

“However, SMILE only runs single-threaded out of the box. While this is not a problem with a small data set, large databases such as TUSZ can take several days to complete. The solution is to utilize a high-performance compute cluster, like the NEDC NeuroNix server [8], in tandem with a workload manager, like Slurm [9]. Through a compute cluster and a workload manager, the SMILE wrapper can be run across the hundreds of CPU cores on the NEDC NeuroNix server, completing the database in only a few hours. The TUSZ database was split into slices containing about 25 EDF files, with each slice running as its own, independent process. To make this work, SMILE had to be modified so that its intermediary file locations were no longer hard coded, with each EDF slice storing its output in a unique directory. From here, the slices are regrouped and scored as development, evaluation, and training datasets.”

“Define evaluation and development data”

A new paragraph was added to discuss the roles of the TUSZ training, development, and evaluation sets.

“A performance assessment of SMILE was conducted using TUSZ's evaluation and development datasets. TUSZ [3] is divided into three distinct datasets, namely, evaluation, development, and training based on an attempt to balance a number of demographic and metadata features of the corpus. The training dataset is typically used to adjust parameters to optimize performance on the development dataset. The evaluation set is treated as a blind dataset and is exclusively employed for model accuracy evaluation. Parameters are not adjusted to maximize performance on the evaluation set.”

“Analyse the datasets used to train the SPaRCNet and NEDC ResNet-18 and compare with the TUSZ.”

This paragraph has been rewritten to address this.

“I believe TUSZ has different montage systems (such as unipolar or bipolar) as well as different subgroups (AR, LE, AR A). which system has been used and why?”

We reworked this paragraph:

“Before the SMILE system can evaluate TUSZ data, the channels that SMILE EDFs use must be appropriately mapped to channels that TUSZ uses. The TUH Corpus contains over $40$ unique channel combinations and four different electrode configurations [10]. SMILE’s lookup table found in the “channel\_mappings.mat” file was modified to include channel labels found in TUSZ data. These labels are mapped in SMILE’s preprocess step which prepares TUSZ data for use of SMILE’s custom montage.”

to address this concern.

“Define LPD, GPD, LRDA, GRDA?”

The first paragraph was modified to reference J. Jing et al and their paper which they describe the seizure classes.

“Seizures of less than 2 seconds has been removed from the analysis. This maybe an absence seizure which young children are more affected. This analysis can only be effective if you have removed the data for young children from the work.”

The following paragraph was added to address this concern:

“Before delving into the outcomes of the SMILE experiments, it is appropriate to comment on the exclusion of seizures lasting less than $2s$. The significance of seizures lasting a short duration remains a topic of debate within the field of neurology, often with the consensus that short seizures primarily affect young children rather than adults. The decision to omit short duration seizures was made as a strategic compromise between sensitivity and FP rates. Inclusion of seizures under $2s$ substantially inflates the false positive rate while offering minimal improvement in sensitivity, primarily due to the scarcity of children experiencing such short seizures in the TUSZ dataset. For example, it was observed that no children aged $0-2$ in TUSZ had encountered seizures lasting less than $2s$. While the precise threshold for defining a “short” seizure can be debated, excluding very brief seizures is a fact of life with most machine learning approaches since they tend to produce a significant number of FPs.”